

ОРИГИНАЛЬНАЯ СТАТЬЯ



DOI: 10.26794/2587-5671-2025-29-6-243-268
 УДК 338.467,339.13,330.131.5(045)
 JEL C51, C53, L82, E22, G11

Прогнозирование финансовой эффективности российского кинематографа с помощью многофакторной ансамблевой модели машинного обучения, тренированной на данных прошлых периодов

А.В. Дождиков

Институт социально-политических исследований ФСНИЦ РАН, Москва, Российская Федерация

АННОТАЦИЯ

Объект исследования – данные проката российского кинематографа с июля 2022 по сентябрь 2023 г., 185 кинокартин. **Инструмент** исследования – предобученные на базе прошлых периодов (с 2004 по июль 2022 г., 1500 кинокартин) 26- и 146-факторные модели машинного обучения. **Цель** исследования – доказать, что модели машинного обучения, обученные на данных прошлых периодов, могут прогнозировать будущие данные. Это особенно **важно** для финансирования программ развития национального кинематографа в России и привлечения частных инвестиций в условиях ухода иностранных кинодистрибуторов с рынка. В исследовании использовались **методы** оценки кинопроектов на основе исторической доходности по прокатным показателям и характеристикам творческих (съемочных) групп. Акцент сделан на ансамблевых моделях – AdaBoost, Bagging, ExtraTrees, GradientBoosting, RandomForest, Stacking, Voting, XGBoost, CatBoost. **Новизна** исследования заключается во введении в научный оборот новых источников и возможности практического применения разработанных подходов для государственного и частного инвесторов при оценке проектов фильмов до начала производственного цикла. **Выводы:** поскольку падение метрик качества (accuracy, гос_aic и других) на выборке из 185 новых кинофильмов (по сентябрь 2023 г.) оказалось незначительным, это открывает возможность использования предобученных моделей на данных прошлых периодов для прогнозирования сборов и других результатов кинопроката. Проанализировав прошлые проекты режиссеров, сценаристов, оператора, продюсеров, художника, монтажера, композитора и ключевых актеров, а также предполагаемые прокатные данные и объем финансирования, можно с высокой точностью предсказать успех фильма. Это позволит оценить общие сборы, окупаемость, количество просмотров и зрительский рейтинг.

Ключевые слова: финансовая эффективность; прогноз киносборов; прокатные данные; национальный кинематограф; машинное обучение; ансамблевые модели; классификация; регрессия; CatBoost; XGBoost; государственная политика в кино; экспорт киноконтента

Для цитирования: Дождиков А.В. Прогнозирование финансовой эффективности российского кинематографа с помощью многофакторной ансамблевой модели машинного обучения, тренированной на данных прошлых периодов. *Финансы: теория и практика*. 2025;29(6):243-268. DOI: 10.26794/2587-5671-2025-29-6-243-268

ORIGINAL PAPER

Forecasting the Financial Efficiency of Russian Cinema Using a Multifactor Ensemble Machine Learning Model Trained on Historical Data

A.V. Dozhdikov

Institute of Socio-Political Studies of the Russian Academy of Sciences, Moscow, Russian Federation

ABSTRACT

The **object** of the study is data on the distribution of Russian cinema films from July 2022 to September 2023. Specifically, it analyzes 185 films that were released during this period. The research tool consists of 26 and 146-factor machine learning models that have been pre-trained based on previous periods (from 2004 to July 2022, with 1,500 films). The **purpose** of the study is to demonstrate that machine learning models, trained on historical data, can accurately predict future data, which is especially important for funding programs aimed at developing national cinema in the Russian Federation and attracting private investment, in light of the departure of foreign film distributors from the film market. The study used **methods** to

© Дождиков А.В., 2025

evaluate film projects based on their historical profitability using rental indicators and the characteristics of the creative teams involved in producing them. The emphasis is on ensemble models – AdaBoost, Bagging, ExtraTrees, GradientBoosting, RandomForest, Stacking, Voting, XGBoost, CatBoost. **The novelty** of this research lies in introducing new sources into the scientific community and the potential for practical application of the developed methods for both public and private investors to evaluate film projects prior to the start of the production cycle. **Conclusions:** Conclusions: Based on the analysis of the quality metrics (accuracy, ROC AUC, and others) for a sample of 185 newly released films (through September 2023), we found that the drop in these metrics was not significant. This suggests that it is possible to use pre-trained models based on historical data to make predictions about fees and other rental outcomes. By analyzing the past work of the project director, screenwriters, cameramen, producers, artists, editor, composer and key actors of the project, estimated distribution data, and the amount of project funding, it is possible to make an accurate prediction about the success of a film. This will allow you to see the total fees, payback period, number of views, and viewer rating.

Keywords: financial efficiency; film box office forecast; box office data; national cinema; machine learning; ensemble models; classification; regression; CatBoost; XGBoost; government policy in cinema; export of film content

For citation: Dozhdikov A.V. Forecasting the financial efficiency of Russian cinema using a multifactor ensemble machine learning model trained on historical data. *Finance: Theory and Practice*. 2025;29(6):243-268. DOI: 10.26794/2587-5671-2025-29-6-243-268

ВВЕДЕНИЕ

На этапе разработки (препродакшен) инвесторы оценивают от нескольких десятков сценариев до нескольких сотен и тысяч в крупных студиях, подбирают ключевых актеров, формируют творческую и съемочную группы и определяют бюджет. Именно на этом этапе желательно рассматривать различные сценарии и комбинации творческих и съемочных групп, прежде чем вкладывать крупную сумму денег в производство даже не одного фильма, а тщательно сформированного и просчитанного «портфеля» кинопроектов. Точный прогноз доходов на этой ранней стадии многократно снижает инвестиционный риск [1]. Огромные первоначальные вложения, связанные с киноиндустрией, как государственные, так и частные, требуют, чтобы инвестиции основывались на надежных и воспроизводимых методах прогнозирования [2], при том, что само кинопроизводство во всем мире пронизано субъективностью и неопределенностью: каждое принятное решение может повлиять как на качество проекта, так и на финансовую сторону проката [3].

Точный прогноз успеха фильма возможен с использованием анализа исторических и текущих данных. Успешность кинопроекта определяется совокупным вкладом актеров, режиссера, композитора, продюсера и других членов креативной команды. Доступ к биографическим данным этих специалистов можно получить из открытых источников, таких как Internet Movie Database (IMDb) [4], Rotten Tomatoes, а также из социальных сетей, где представлена информация о подписчиках этих деятелей искусства. Кроме того, можно ознакомиться с данными о количестве просмотров проектов в различных базах данных и аналитическими материалами, включая трейлеры и другие релевантные материалы [5].

Прогнозировать можно не только количественные данные финансового успеха и число просмотров [6],

но и результаты кинопремий «Оскар» [7] и другие «качественные» события.

Прогнозирование кинопроката и кассовых сборов с помощью интеллектуального анализа данных развито в Голливуде и индийском Болливуде [8], равно как и в нигерийском Нолливуде [9]. Азиатский тренд — активное использование машинного обучения и нейросетей. Большинство научных публикаций по этой тематике выходит в материковом Китае, Корее, Тайване, Индии, Индонезии, Шри-Ланке и других странах Азии.

Прогнозирование кассовых сборов, как правило, осуществляется с использованием алгоритмов машинного обучения [10], в том числе с использованием больших данных перед непосредственным выпуском проектов в широкий прокат, потому что результаты проката напрямую влияют на инвестиционные решения на фондовом рынке, разработку стратегий продвижения рекламных компаний, планирование показов фильмов кинотеатральными сетями [11]. Тем не менее возможно прогнозирование результатов и на ранней стадии производства (препродакшен), когда съемки самого проекта еще не стартовали. Прогнозирование прибыльности фильмов на ранней стадии производства может быть полезным для принятия решения об инвестировании [12], о выпуске акций и облигаций и использовании иных финансовых инструментов.

В России индустрия прогнозирования национального кинопроката развита в меньшей мере, чем, например, в Китае. Однако доказанная в работе [13] предсказуемость российского кинематографа, наличие четких трендов по окупаемости, просмотрам, оценкам зрителей в зависимости от характеристик и творческого состава киносъемочных групп позволит более успешно реализовать государственную политику в области развития искусства, культуры и креативных индустрий. Все прогнозы и выводы могут осуществляться до начала самого затратного процесса съемок, а базовые данные по факторам «успешности», включая

длительность кинопроекта, жанровую направленность, возрастной рейтинг и другие элементы уже определены автором в его предшествовавшей научной работе [14].

МЕТОДЫ ИССЛЕДОВАНИЯ И ИСТОЧНИКИ ДАННЫХ

Прогнозирование кинопроката прошло долгий этап от статистического моделирования начала 2000-х гг. до методов машинного обучения второй половины 2010-х гг. [15]. Как уже было отмечено, точный прогноз валовых кассовых сборов приносит огромную пользу для инвестиций и управления киноиндустрией, соответственно, применение разных методов прогнозирования и набора экономических факторов позволит выявить наиболее продуктивные сочетания [16].

В большинстве исследований используется многофакторное или мультимодальное прогнозирование рейтинга фильмов на основе открытых данных [17] с числом факторов от нескольких до нескольких сотен. Используются как данные локального кинорынка, например тайваньского [18], так и глобальные базы данных, как IMDB [19].

Изучаться могут как сами прокатные данные и их история, так и сопутствующие, например поисковые запросы и тенденции [20], может осуществляться корреляционный анализ сетевых больших данных и данных временных рядов фильмов [21]. Существуют также текстовые методики, включающие использование языковых моделей для анализа текста комментариев пользователей в отношении будущего проекта, соответственно, возможен частотный анализ слов в рецензиях [22] и поиск их сочетаний, прогнозирующих успех.

В настоящей работе **объект исследований** ограничен, изучаются именно прокатные исторические данные, прогнозируется успех/неуспех фильма и более сложная классификация, количество просмотров, сборы, окупаемость (соотношение сборы/бюджет) и зрительский рейтинг.

Как правило, исследователи сравнивают различные методы машинного обучения на предмет эффективности и точности [23], поскольку заранее невозможно спрогнозировать результаты того или иного метода без детального анализа исходного массива данных.

Прогнозирование успеха фильма с помощью методов машинного обучения требует использования различных способов повышения точности [24]. В настоящем исследовании этот вопрос решается с помощью настройки гиперпараметров машинного обучения, а также использования синтетических данных для балансировки предсказываемых классов.

Как правило, исследование начинается с линейной регрессии, логистической регрессии иди «машины опорных векторов» [25]. Данный промежуточный этап

был представлен в прошлой научной работе [14]. Отметим, что похожий подход был использован при анализе национального кинематографа Турции, включая множественный регрессионный анализ, гребневую регрессию и лasso-регрессию, древовидные методы, такие как «случайный лес», SVM и KNN и наилучшие результаты вполне предсказуемо были получены на базе метода «случайный лес» [26].

В первую очередь исследователя должны интересовать так называемые ансамблевые методы, дающие при правильной настройке более высокую точность классификации, позволяющие использовать многофакторные модели, такие как XGBoost, LightGBM, CatBoost, которые применяли коллеги из Китая при анализе кинопроката с учетом данных о пандемии COVID-19 [27]. Ансамблевые методы на основе деревьев решений эффективнее, чем методы, использующие алгоритмы, похожие на k-NN [28]. Исследователи из Кореи [29] предсказывают, что стекинг станет важным инструментом. Этот метод объединяет все алгоритмы прогнозирования в одну большую метамодель.

Исходный датасет состоит из 1685 записей о национальных фильмах, начиная с фильма «72 метра» (ID_kinopoisk = 70952, дата выхода в прокат — 12.02.2004) по «Концерт отменяется» (ID_kinopoisk = 5325618, дата выхода в прокат — 07.09.2023). Данные собраны автором из открытых источников.

Датасет был разделен на 2 части — 1500 фильмов для обучения и тестирования моделей машинного обучения и контрольный, валидационный датасет на 185 кинокартин в хронологическом порядке выхода. К последнему датасету модели машинного обучения не имели доступа в период обучения. Первый датасет был поделен на тренировочную и обучающую выборки в пропорции 70/30 с фиксированными параметрами разделения (random state = 42 для train_test_split). Категориальные переменные (жанры) были закодированы LabelEncoder, пропуски заполнены медианными значениями.

Примерно 30% российских фильмов, которые идут в прокате, не имеют открытых бюджетов. Это существенно влияет на точность исследований и искажает результаты регрессионных моделей.

Гипотеза исследования связана с тем, что модели, обученные и протестированные на данных прошлых периодов (при случайному распределении проектов для тестовой и тренировочной выборки), покажут достаточно высокие метрики и показатели качества даже в случае использования контрольного датасета из данных новых проектов в хронологическом порядке в узких временных рамках с учетом объективных изменений в структуре проката, вызванных уходом иностранных кинодистрибутеров. Таким образом, предобученные модели будут работать в заведомо худших услови-

ях с точки зрения изменения рыночных факторов и «качества» данных. Если на контрольном датасете уменьшение метрик качества будет незначительным по сравнению с тестовыми данными, то основной тезис исследования будет доказан. Многофакторные модели, обученные на исторических данных прошлых периодов, эффективны. На практике, обучив модель на данных 2004–2024 гг., мы достаточно точно сможем спрогнозировать судьбу всех национальных проектов в 2025 и 2026 гг. и в другие периоды времени.

ПРЕДВАРИТЕЛЬНЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Для пилотного подхода использована упрощенная версия, 26-факторная модель, включавшая следующие данные: неделю, месяц выхода фильма, количество экранов, а также такие позиции, как бюджет фильма (в млн руб.), возрастной рейтинг фильма от 0+ до 18+, общая длительность фильма в минутах, среднее соотношение сборы/бюджет по жанру, средний зрительский рейтинг Кинопоиска по жанру, средние сборы по жанру, среднее количество просмотров для жанра, соотношение сборы/бюджет для поджанра проекта, средний зрительский рейтинг для поджанра проекта, средние сборы для поджанра проекта, средние просмотры (количество) для поджанра проекта, среднее соотношение сборы/бюджет для проектов режиссера, средний рейтинг режиссера на Кинопоиске, средняя величина сборов проектов данного режиссера, средняя величина просмотров фильмов режиссера, среднее соотношение сборы/бюджет для сценариста, средний зрительский рейтинг для сценариста, средняя величина сборов по проектам сценариста, среднее количество просмотров проектов сценариста, соотношение сборы/бюджет второго сценариста, средний зрительский рейтинг второго сценариста, среднее количество сборов проектов второго сценариста, среднее количество просмотров проектов второго сценариста. Таким образом, пилотный датасет ограничен общими прокатными данными и историческими данными режиссеров и сценаристов (в случае с режиссерами новичками и сценаристами использованы медианные значения для недопущения их дискриминации).

Как правило, для данного типа задач может использоваться бинарная классификация: проект не окупился в прокате, собрал менее 2-х своих бюджетов с классом — 0, и проект окупился в прокате, собрав 2 и более своих бюджетов, — 1. Однако есть возможность с незначительной потерей точности прогнозировать и более сложный класс проекта, например: 0 — провалился в прокате, 1 — не окупился, 2 — окупился, 3 — принес прибыль свыше 100% производственного бюджета. На следующем этапе регрессии будут определены сборы

(box), просмотры (views), соотношение сборы/бюджет (box_buget) и зрительский рейтинг (kinopoisk_R).

В качестве теста было использовано несколько простых ансамблевых моделей машинного обучения AdaBoostClassifier, BaggingClassifier, ExtraTreesClassifier, GradientBoostingClassifier, RandomForestClassifier, StackingClassifier, VotingClassifier, HistGradientBoostingClassifier. Для четырехклассовой классификации наилучшие результаты по параметрам accuracy¹ = 0,8878 и ROC_AUC score² = 0,9611 показала модель HistGradientBoostingClassifier. С подобранными методом GridSearchCV гиперпараметрами (learning_rate = 0,1, max_depth = 4, max_iter = 80) получена точность = 0,8918 и ROC_AUC score = 0,9653. Подробнее — на рис. 1.

Вслед за исследователями из Китая отметим, что XGBoosts и ему подобные модели являются эффективными ансамблевыми алгоритмами для решения подобного класса задач [30].

При переводе меток классов в формате one-vs-all получены ROC_AUC score для каждого класса (рис. 2).

Определим с помощью 26-факторной модели количественный показатель, зрительский рейтинг Кинопоиска (безразмерный коэффициент). Воспользуемся перебором нескольких вариантов моделей машинного обучения: AdaBoostRegressor, HistGradientBoostingRegressor, BaggingRegressor, ExtraTreesRegressor, GradientBoostingRegressor, RandomForestRegressor, CatBoostRegressor. В качестве оценочных метрик каждой модели используем стандартный набор MSE (Mean Squared Error) — средняя квадратичная ошибка, R 2 (R-квадрат) — коэффициент детерминации, который оценивает прогностическую способность модели, доля объясненной дисперсии относительно общей дисперсии и MAE — (Mean Absolute Error) — средняя абсолютная ошибка (см. таблицу).

Лучшие результаты в данном случае показала ансамблевая модель ExtraTreesRegressor с подобранными гиперпараметрами (max_depth = 10, min_samples_split = 10, n_estimators = 250), ее показатели улучшены до MSE = 0,2356, MAE = 0,3010, R 2 = 0,8492. Таким образом даже простая 26-факторная модель может давать количественные прогнозы в отношении зрительских рейтингов.

¹ Accuracy (точность) — количество правильно предсказанных результатов классификации ко всему количеству предсказанных результатов, колеблется от 0 до 1.

² Кривая ROC (Receiver Operating Characteristic или рабочая характеристика приемника) — это график, иллюстрирующий производительность классификационной модели при всех возможных порогах классификации. Показатель AUC (Area Under the ROC Curve) — это мера, которая позволяет суммировать производительность модели одним числом, измеряя площадь под кривой ROC. AUC колеблется от 0 до 1, где более высокое значение AUC указывает на более высокую производительность модели. Показатель в 0,5 соответствует «случайному гаданию».

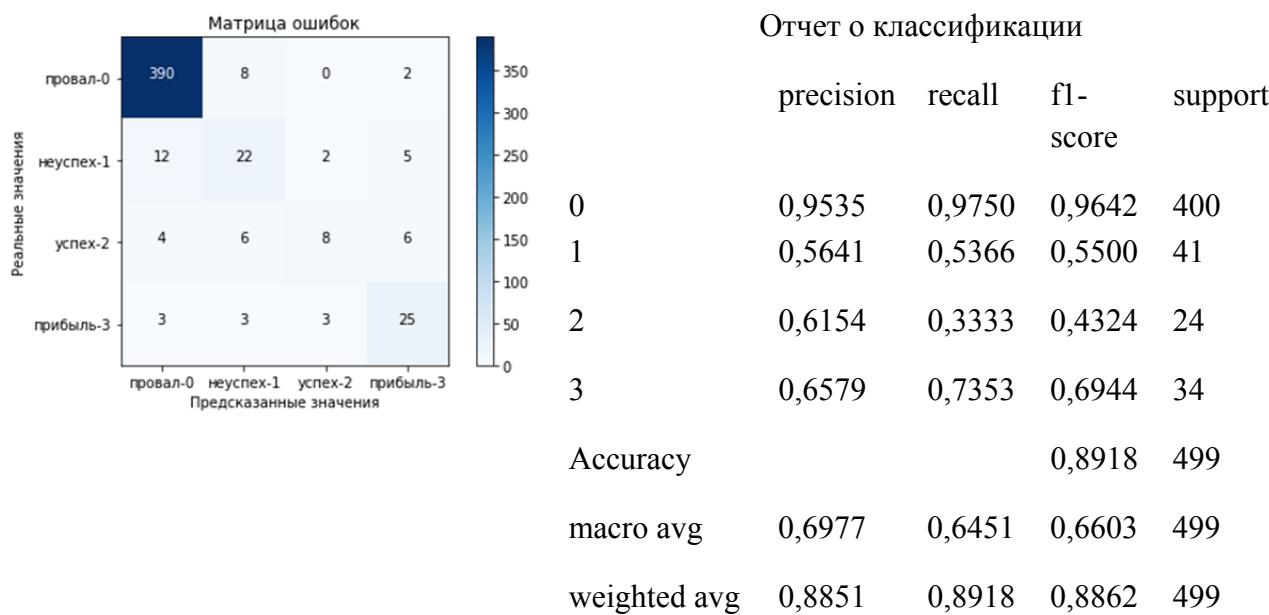


Рис. 1 / Fig. 1. Матрица ошибок и отчет о классификации 26-факторной модели / Error Matrix and Classification Report of the 26-Factor Model

Источник / Source: составлено автором / Compiled by the author.

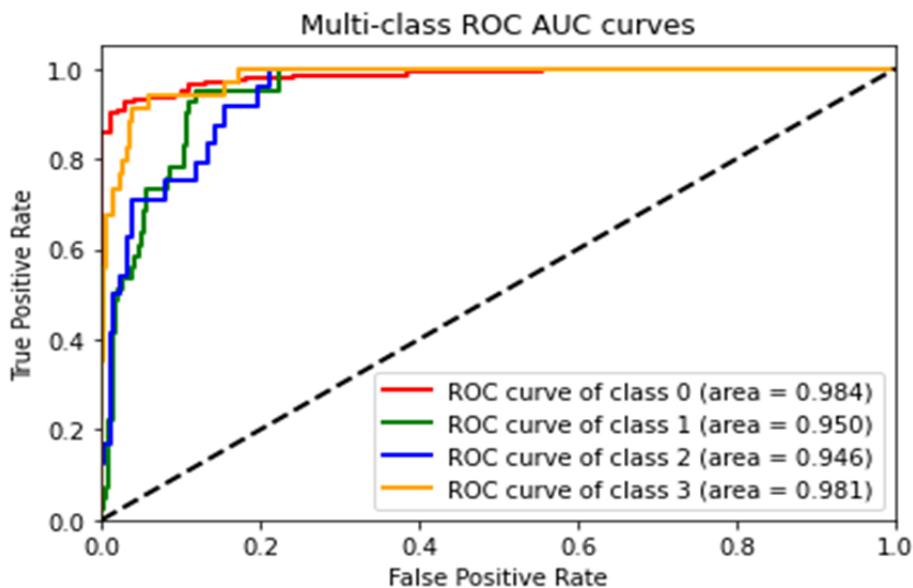


Рис. 2 / Fig. 2. Кривая ROC_AUC_score для четырехклассовой 26-факторной модели / ROC_AUC_Score Curve for a Four-Class 26-Factor Model

Источник / Source: составлено автором / Compiled by the author.

В настоящей работе с учетом ограничений по объему приводятся только общие данные по разработанной 26- и 146-фактороной регрессионным моделям, предсказывающим абсолютные значения сборов, количества просмотров, окупаемости и зрительского рейтинга.

МНОГОФАКТОРНАЯ (ПОЛНАЯ) МОДЕЛЬ КЛАССИФИКАЦИИ

Для доказательства гипотезы использована более сложная прогностическая модель, учитывающая как

данные проката прошлых периодов, так и расширенный перечень участников творческой группы. 146-параметрическая модель, включающая следующие показатели: неделя и месяц выхода фильма в прокат, количество экранов проката, бюджет фильма (где известен), возрастной рейтинг, длительность, жанры (до 5 наименований), исторические данные по режиссеру проекта, сценаристам, оператору, продюсерам проекта, художнику, монтажеру, композитору, основным актерам проекта. В перечень истори-

Таблица / Table

Характеристики нескольких прогностических ансамблевых моделей регрессии для рейтинга Кинопоиска / Characteristics of Several Predictive Ensemble Regression Models for Kinopoisk Ratings

№	Модель машинного обучения / Machine learning model	MSE	R 2	MAE
1	AdaBoostRegressor	0,385185	0,753436	0,457239
2	BaggingRegressor	0,283048	0,818815	0,332489
3	ExtraTreesRegressor	0,238982	0,847023	0,301738
4	GradientBoostingRegressor	0,270861	0,826616	0,338732
5	RandomForestRegressor	0,255361	0,836538	0,306668
6	HistGradientBoostingRegressor	0,294991	0,811170	0,351616
7	CatBoostRegressor	0,261589	0,832552	0,337352

Источник / Source: составлено автором / Compiled by the author.

ческих данных включались усредненные исторические данные по экранам проката, бюджетам, сборам, просмотрам, рейтингу Кинопоиска, IMDb (при наличии), относительные показатели — сборы на экран, просмотры на экран, сборы к бюджету.

Для доказательства исходного утверждения общий датасет из 1685 фильмов разделен на 3 части. На 1500 записи (в выборках записи распределены случайным образом) модель машинного обучения тренируется и тестируется, на 185, расположенных уже в хронологическом порядке с момента выхода в середине 2022 по сентябрь 2023 г., — проверяется обобщающая способность модели.

Уже для двухклассовой классификации актуальна проблема дисбаланса классов: количество фильмов, успешных в прокате (класс = 1), на порядок меньше неудачных (класс = 0). Поэтому целесообразно использование синтетических данных с использованием библиотеки imbalanced-learn. Таким образом, общее количество записей из основного датасета в 1500 становится 2670. Выборка разделяется на тренировочную и тестовую в соотношении 70 к 30 с помощью train_test_split. По аналогии с 26-факторной моделью подобран алгоритм машинного обучения, дающий в данной конфигурации наилучшие метрики качества. Им оказалась отечественная разработка от специалистов «Яндекса»⁵ CatBoostClassifier.

Обученная с помощью синтетических данных на сбалансированных классах модель ошиблась только

в 12 случаях, показав точность в 0,9863 и предельный ROC_AUC_score = 0,9999 после 1000 итераций. В принципе это может свидетельствовать о «переобученности» модели и, соответственно, о низкой прогностической способности. Для проверки воспользуемся фактическими данными по 185 фильмам, к которым модель по условиям проверки гипотезы не имела доступа ни на этапе обучения, ни на этапе тестирования.

Несмотря на предположение о переобученности модели, несбалансированность валидационных данных по классам (170 — неуспешных проектов, 15 — окупившихся в прокате) и тот факт, что они отражают новый исторический период развития российского кинематографа, связанный с уходом иностранных дистрибутеров, классифицирующая способность модели осталась крайне высокой. Точность = 0,9568, ROC_AUC_score = 0,9803.

Аналогично используем новые данные для проверки модели. В силу специфики данных результаты оказались незначительно ниже тестовых, что опять же не исключает возможности использования подобного подхода для прогнозирования данных новых кинофильмов на стадии препродакшен. Точность четырехклассовой классификации составила 0,9189, ROC_AUC_score = 0,9663.

При оценке метрик precision, recall, f1-score и ROC_AUC отметим, что самые высокие результаты — у крайних классов (0 и 3), а больше всего ошибок в отношении идентификации промежуточных классов 1 и 2. Это связано, предположительно, с недостатком информации по бюджету проектов (до 30%) и условными критериями окупаемости: два бюджета проекта — это, скорее всего, эмпирическое правило для определения окупаемости проекта.

⁵ CatBoost — это быстрая, масштабируемая и высокопроизводительная библиотека повышения градиента деревьев решений. Используется для ранжирования, классификации, регрессии и других задач машинного обучения. URL: <https://github.com/catboost/> (дата обращения: 15.03.2024).

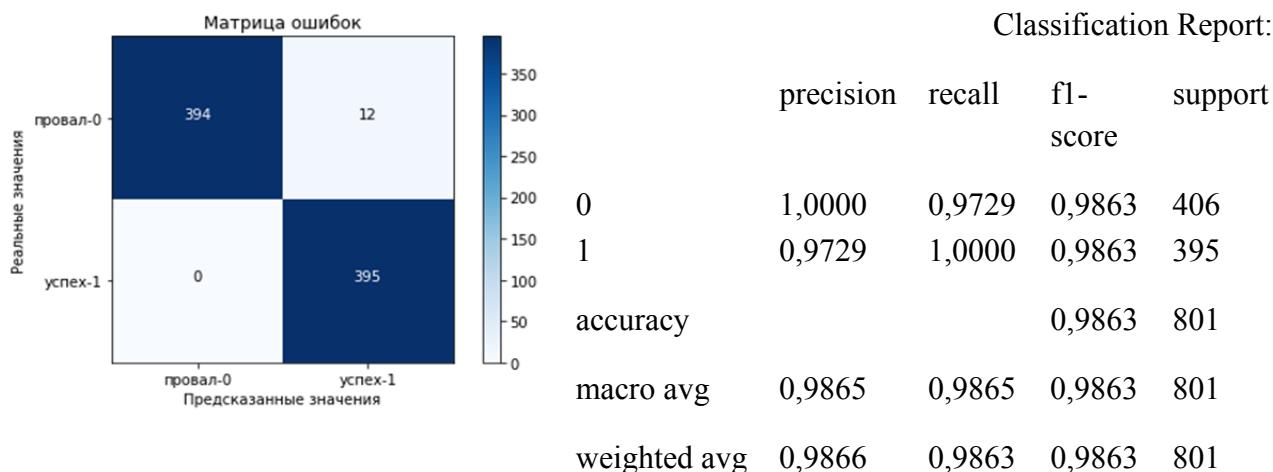


Рис. 3 / Fig. 3. Матрица ошибок и отчет о классификации 146-факторной модели / Error Matrix and Classification Report of the 146-Factor Model

Источник / Source: составлено автором / Compiled by the author.

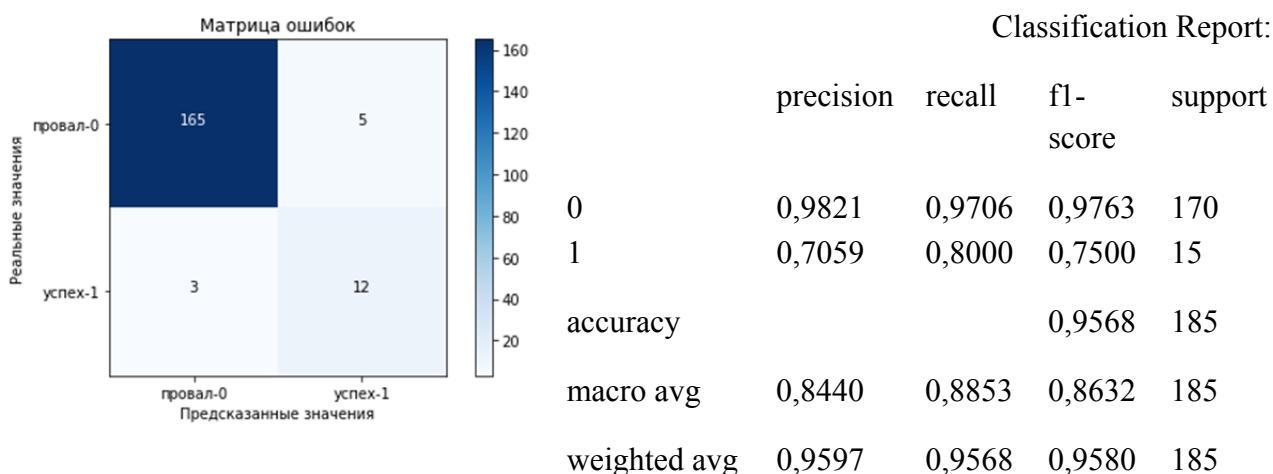


Рис. 4 / Fig. 4. Матрица ошибок и отчет о классификации 146-факторной модели на новых данных / Error Matrix and Classification Report of the 146-Factor Model based on New Data

Источник / Source: составлено автором / Compiled by the author.

По результатам эксперимента доказана гипотеза о предсказуемости финансовых результатов российского кинематографа с помощью модели машинного обучения, обученной на данных прошлых периодов, для новых фильмов, в том числе находящихся в предварительной стадии препродакшен, до начала самого затратного этапа съемок. Практические примеры по каждому российскому фильму (его предсказанный класс и фактический) с учетом ошибок приведены в Приложении А — в варианте двухклассовой и четырехклассовой классификации.

Некоторое уменьшение метрик классификации может быть обусловлено: «переобучением» модели, использованием синтетических данных для выравнивания дисбаланса по классам, хронологически «узким»

периодом выхода 185 кинокартин, их подачей в хронологическом порядке, отчасти — фундаментальными изменениями характеристик российского кинорынка, обусловленными уходом иностранных дистрибутеров. Каждый эпизод из данных вариантов нуждается в дополнительном исследовании. Тем не менее, с учетом возможного дообучения прогностических моделей, на новых данных получена возможность быстрой предварительной оценки перспектив как отдельного кинопроекта, так и группы кинопроектов с точки зрения их будущих финансовых и иных прокатных результатов.

Продолжение исследования связано с получением более точных данных по бюджетам (отмечен дефицит данных по бюджету в 30% российских фильмов, что негативно влияет на точность прогнозирования).

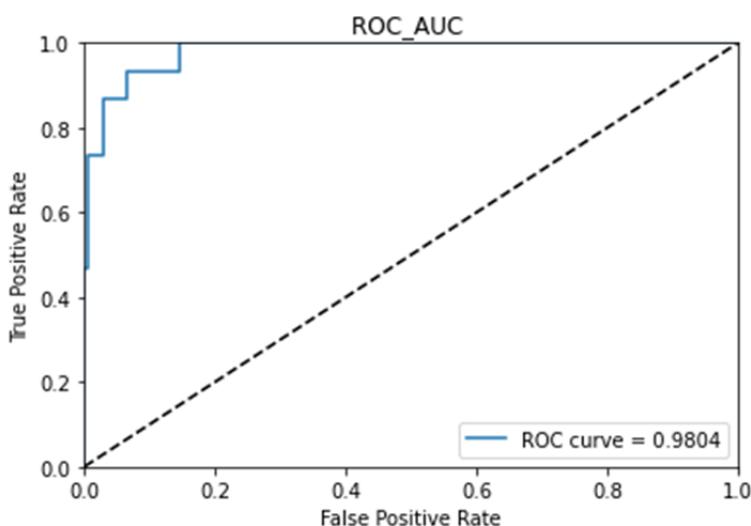


Рис. 5 / Fig. 5. Кривая ROC_AUC_score для двухклассовой 146-факторной модели на новых данных / ROC_AUC_Score Curve for a Two-Class 146-Factor Model on New Data

Источник / Source: составлено автором / Compiled by the author.

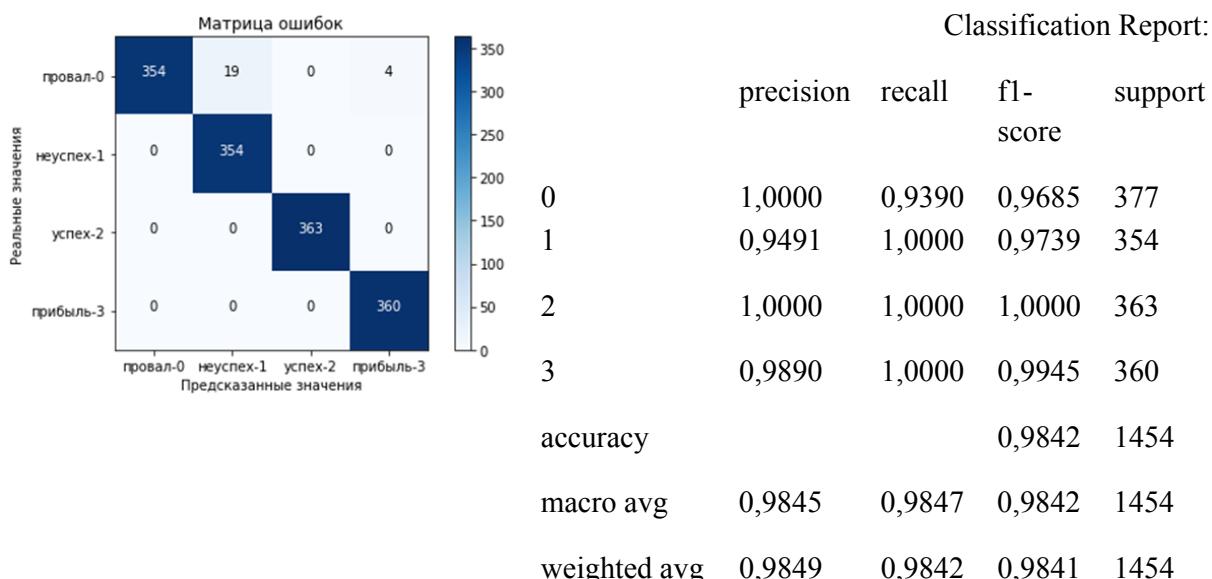


Рис. 6 / Fig. 6. Матрица ошибок и отчет о классификации 146-факторной модели (4 класса) / Error Matrix and Classification Report of the 146-Factor Model (4 Classes)

Источник / Source: составлено автором / Compiled by the author.

Низкая информационная открытость бухгалтерии российского кинопроизводства отрицательно сказывается на возможности точного прогнозирования сборов на основании регрессионных моделей машинного обучения. Тем не менее возможность таких прогнозов есть.

185 фильмов — это 21,7 млрд руб. сборов и 73,5 млн просмотров. Основной фактор — аномалия «Чебурашки»⁴, проект собрал свыше 7 млрд руб. и 23,5 млн

просмотров. Данный фильм является аномалией, обусловленной выходом в новогодний период, отсутствием конкурентов и объективным фактором — редким сочетанием по творческой группе (режиссер с самой большой в истории серией коммерчески удачных работ, актеры и другие участники творческой группы с высокими прокатными показателями). С учетом

индустрис! Плас. 21.03.2023. URL: <https://plusworld.ru/journal/2023/plus-3-2023/uspekh-cheburashki-mozhno-povtorit-s-pomoshchyu-ii-i-ne-tolko-v-kinoindustrii/> (дата обращения: 15.03.2024).

⁴ См. публикацию: Дождиков А.В. Успех «Чебурашки» можно повторить с помощью ИИ. И не только в кино-

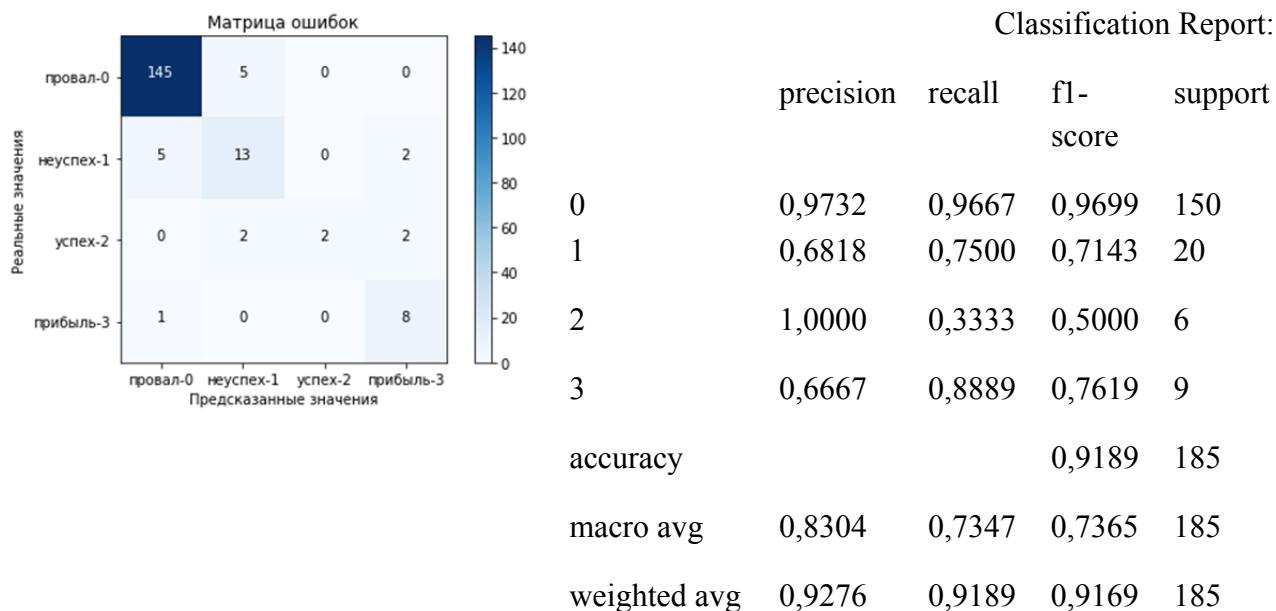


Рис. 7 / Fig. 7. Матрица ошибок и отчет о классификации 146-факторной модели (4 класса) на новых данных / Error Matrix and Classification Report of the 146-Factor Model (4 Classes) Based on New Data
Источник / Source: составлено автором / Compiled by the author.

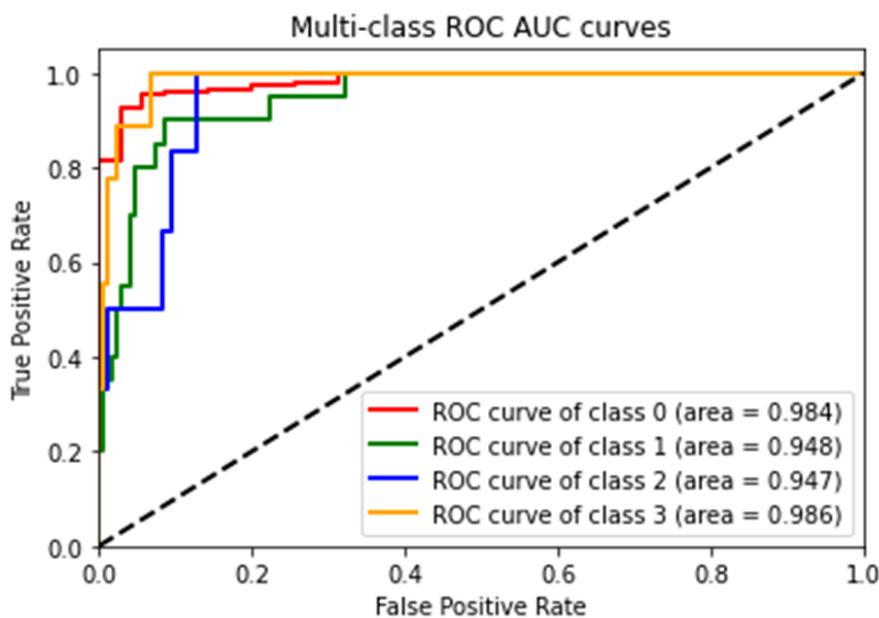


Рис. 8 / Fig. 8. Кривая ROC_AUC_score для четырехклассовой 146-факторной модели на новых данных / ROC_AUC_Score Curve for a Four-Class 146-Factor Model on New Data

Источник / Source: составлено автором / Compiled by the author.

исключения результатов по «Чебурашке» и отсутствующим данным по рейтингу Кинопоиска и бюджетам получили точность прогнозирования (R² коэффициент детерминации) для фактических и предсказанных значений: зрительский рейтинг Кинопоиска — 0,8392, сборы в прокате — 0,9076, просмотры — 0,8797. Средняя абсолютная ошибка MAE (mean absolute error) для сборов составила 26,78 млн руб., для просмотров — 108,23 тыс., для рейтинга Кинопоиска — 0,5874 пунктов рейтинга. Таким образом, по величине средней абсо-

лютной ошибки данная методика прогнозирования не подходит для малобюджетных (как правило, авторских или документальных) картин с небольшим числом экранов в прокате по сборам и просмотрам. Рекомендуется оценивать в первую очередь среднебюджетные (80–300 млн руб.) и крупнобюджетные (до 850 млн руб.) фильмы. Проекты с большим бюджетом, как правило, не окупаются в прокате и не собирают двух своих бюджетов в связи с ограниченной емкостью российского рынка.



Рис. 9 / Fig. 9. Расхождение фактических и предсказанных значений сборов, просмотров и зрительского рейтинга / Discrepancy Between Actual and Predicted Values of Box Office, Views and Viewership Ratings

Источник / Source: составлено автором / Compiled by the author.

Подробные данные по фактическим и предсказанным значениям — см. *Приложения А и Б*.

Индивидуальные результаты — предсказанные и фактические могут отличаться друг от друга для фильмов с минимальными и максимальными показателями. Однако для «портфеля кинопроектов» расхождение между предсказанными показателями и фактическими будет существенно меньше. Чем больше пул оцениваемых проектов, тем меньше расхождение между суммарным результатом между предсказанными и фактическими данными по финансовым результатам.

НАПРАВЛЕНИЯ ИССЛЕДОВАНИЯ

Дальнейшее развитие методик прогнозирования, как зрительского рейтинга [31], так и кассовых сборов [32], предполагает использование нейросетей, в том числе «глубоких сетей доверия» — Deep Belief Network или DBN [33].

Возможно комбинирование подходов классического машинного обучения и нейронных сетей типа IFOA-GRNN [34]. Также перспективной представляется комбинация классических методов машинного обучения типа «машины опорных векторов», нейросетей и методов обработки естественного языка в сочетании с традиционными историческими данными из источников IMDb, Rotten Tomatoes, Box Office Mojo и Metacritic [27].

Отдельное направление, предполагающее задействование элементов «компьютерного зрения», — использование графических изображений, включая постеры фильма, анализируемые с помощью извлечения признаков для сверточной нейронной сети. Использование алгоритмов глубокого обучения позволит создать системы поддержки принятия решений нового поколения для кинопроизводства, а также перспективные «гибридные» человеко-машинные системы принятия решений, основанные на взаимодействии человека и машины.

ВЫВОДЫ

В *Приложениях А и Б* приведены результаты классификации по 185 кинокартинам (двухклассовая и четырехклассовая классификации) и результаты прогнозирования количественных данных по 185 кинокартинам. Указаны прогнозные и фактические значения.

Отметим высокую точность результатов именно классификации. Прогностическая способность регрессных моделей (сколько именно заработает данный фильм и сколько зрителей его просмотрит) ограничена следующими факторами:

1. 30% бюджетов российских кинопроектов неизвестны (скрыты продюсерами от общественности).

2. На итоговые результаты влияет ряд дополнительных параметров, в частности, маркетинговый бюджет, характеристики рекламной кампании, время выхода фильма в прокат, которое может быть изменено «административным» решением, конкуренция с другими картинами, каннибализация целевой аудитории.

3. Более или менее точные количественные прогнозы возможны в отношении большинства средне-бюджетных и крупнобюджетных картин. Совпадение прогнозных результатов с фактическими меньше всего у проектов-блокбастеров, в прокатную судьбу которых включается «административный» ресурс и у малобюджетных проектов.

С учетом этих данных и дообучения моделей прогнозирования на данных новых периодов появляются возможности:

- оценки проектов кинофильмов на этапе продюсерского питчинга (Фонд Кино, Министерство культуры, частные организации) до начала съемочного периода; для получения предварительного прогноза достаточно предполагаемых прокатных характеристик фильма, сведений о творческой группе;

- формирования портфеля кинопроектов для привлечения частно-государственных инвестиций; 20–30 кинофильмов в «портфеле» в год предполагают стабильную доходность на среднесрочном горизонте планирования;

- создания долгосрочных стратегий развития национального кинематографа;

- разработки нормативных стратегических документов в сфере поддержки национального кинематографа;

- подбора прокатных характеристик и характеристик творческой группы для достижения максимального эффекта по охвату аудитории и зрительским рейтингам при производстве национальных фильмов и проектов с приоритетной государственной поддержкой.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Paul C., Das P.K. Predicting movie revenue before committing significant investments. *Journal of Media Economics*. 2022;34(2):63–90. DOI: 10.1080/08997764.2022.2066108
2. Mbunge E., Fashoto S.G., Bimha H. Prediction of box-office success: A review of trends and machine learning computational models. *International Journal of Business Intelligence and Data Mining*. 2022;20(2):192–207. DOI: 10.1504/IJBDM.2022.120825

3. Tantawichien J., Mizuyama H., Nonaka T. Designing a human computation game for enhancing early-phase movie box office prediction. In: Hamada R., et al. Neo-simulation and gaming toward active learning. Singapore: Springer; 2019:13–22. (Translational Systems Sciences. Vol. 18.). DOI: 10.1007/978–981–13–8039–6_2
4. Pirunthavi V., Vithusia R.P., Abishankar K., Ekanayake E.M., Yanusha M. Movie success and rating prediction using data mining algorithms. In: Int. res. conf. of Uva Wellassa University (IRCUWU-2020) “Sustainable business transition through information and knowledge dissemination”. Badulla: Uva Wellassa University; 2020:175–176. URL: https://www.uwu.ac.lk/wp-content/uploads/2020/proceeding_of_ircuwu2020_v2.pdf
5. Chakraborty P., Rahman Z., Rahman S. Movie success prediction using historical and current data mining. *International Journal of Computer Applications*. 2019;178(47):1–5. DOI: 10.5120/ijca2019919415
6. Murschetz P.C., Bruneel C., Guy J.L., et al. Movie industry economics: How data analytics can help predict movies' financial success. *Nordic Journal of Media Management*. 2020;1(3):339–359. DOI: 10.5278/njmm.2597–0445.5871
7. Bruneel C., Guy J.-L., Haughton D., et al. Movie analytics and the future of film finance. Are Oscars and box office revenue predictable? In: Murschetz P., Teichmann R., Karmasin M., eds. Handbook of state aid for film. Media business and innovation. Cham: Springer; 2018:551–578. DOI: 10.1007/978–3–319–71716–6_30
8. Ruus R., Sharma R. Predicting movies' box office result — a large scale study across Hollywood and Bollywood. In: Cherifi H., Gaito S., Mendes J., Moro E., Rocha L., eds. Complex networks and their applications VIII. (COMPLEX NETWORKS 2019). Cham: Springer; 2020:982–994. (Studies in Computational Intelligence. Vol. 882). DOI: 10.1007/978–3–030–36683–4_78
9. Adekola O.D., Maitanmi S.O., Kasali F.A., et al. Movie success prediction using data mining. *British Journal of Computer, Networking and Information Technology*. 2021;4(2):22–30. DOI: 10.52589/BJCNIT-CQOCIREC
10. Namli Ö., Ozcan T. Forecasting of box office revenue using machine learning algorithms. *Yönetim Bilişim Sistemleri Dergisi*. 2017;(3):130–143.
11. Wang Z., Zhang J., Ji S., et al. Predicting and ranking box office revenue of movies based on Big Data. *Information Fusion*. 2020;60:25–40. DOI: 10.1016/j.inffus.2020.02.002
12. Shahid M.H., Islam M.A. Investigation of time series-based genre popularity features for box office success prediction. *The Open Access Journal for Computer Science*. 2023;9: e1603. DOI: 10.7717/peerj-cs.1603
13. Дождиков А.В. Прогнозирование результатов кинопроката с помощью машинного обучения. *Вопросы теоретической экономики*. 2023(4):93–114. DOI: 10.52342/2587–7666VTE_2023_4_93_114
14. Dozhdkov A. Prediction of the results of movie release using machine learning. *Voprosy teoreticheskoi ekonomiki = Theoretical Economics*. 2023;(4):93–114. (In Russ.). DOI: 10.52342/2587–7666VTE_2023_4_93_114
15. Дождиков А.В. Определение инвестиционного успеха и его факторов для российского кино в прокате с помощью машинного обучения. *Финансы: теория и практика*. 2024;28(1):188–203. DOI: 10.26794/2587–5671–2024–28–1–188–203
16. Dozhdkov A.V. Determination of investment success and its factors for Russian cinema at the box office using machine learning. *Finance: Theory and Practice*. 2024;28(1):188–203. DOI: 10.26794/2587–5671–2024–28–1–188–203
17. Abidi S.M.R., Xu Y., Ni J., Wang X., Zhang W. Popularity prediction of movies: From statistical modeling to machine learning techniques. *Multimedia Tools and Applications*. 2020;79(4):35583–35617. DOI: 10.1007/s11042–019–08546–5
18. Li D., Liu Z.-P. Predicting box-office markets with machine learning methods. *Entropy*. 2022;24(5):711. DOI: 10.3390/e24050711
19. Qin M., Zhou Q., Chen W., Zhao L. MAMRP: Multi-modal data aware movie rating prediction. In: Yang X., et al. Advanced data mining and applications (ADMA 2023). Cham: Springer; 2023:660–675. (Lecture Notes in Computer Science. Vol. 14177). DOI: 10.1007/978–3–031–46664–9_44
20. Lu S.-H., Wang H.-J., Nguyen A.T. Machine learning applications on box-office revenue forecasting: The Taiwanese film market case study. In: Optimal transport statistics for economics and related topics. Cham: Springer; 2023:384–402. (Studies in Systems, Decision and Control. Vol. 483). DOI: 10.1007/978–3–031–35763–3_49
21. Gore M., Sheth A., Abbad S., Jain P., Mishra P. IMDB box office prediction using machine learning algorithms. *International Journal for Research in Applied Science and Engineering Technology*. 2022;10(5):2438–2442. DOI: 10.22214/ijraset.2022.42653
22. Kang D. Box-office forecasting in Korea using search trend data: A modified generalized Bass diffusion model. *Electronic Commerce Research*. 2021;21(3):41–72. DOI: 10.1007/s10660–020–09456–7
23. Li Na., Xia L. Correlation analysis of network Big Data and film time-series data based on machine learning algorithm. *Mathematical Problems in Engineering*. 2022. DOI: 10.1155/2022/4067554

24. Yoo B.-K., Kim S.-H. Movie box office prediction at the distribution stage using text mining of movie reviews. *The Korean Logistics Research Association*. 2023;33(1):95–105. DOI: 10.17825/klr.2023.33.1.95
25. Quader N., Gani O., Chaki D. Performance evaluation of seven machine learning classification techniques for movie box office success prediction. In: 3rd Int. conf. on electrical information and communication technology (EICT). (Khulna, December 7–9, 2017). New York, NY: IEEE; 2017:1–6. DOI: 10.1109/EICT.2017.8275242
26. Lee K., Park J., Kim I., Choi Y. Predicting movie success with machine learning techniques: Ways to improve accuracy. *Information Systems Frontiers*. 2018;20(3):577–588. DOI: 10.1007/s10796-016-9689-z
27. Hu Y.-H., Shiau W.-M., Shih S.-P., Chen C.-J. Considering online consumer reviews to predict movie box-office performance between the years 2009 and 2014 in the US. *The Electronic Library*. 2018;36(6):1010–1026. DOI: 10.1108/EL-02-2018-0040
28. Gürbüz A., Biçer E., Kaya T. Prediction of gross movie revenue in the Turkish box office using machine learning techniques. In: Kahraman C., et al., eds. Intelligent and fuzzy systems (INFUS 2022). Cham: Springer; 2022:86–92. (Lecture Notes in Networks and Systems. Vol. 505). DOI: 10.1007/978-3-031-09176-6_10
29. Ni Y., Dong F., Zou M., Li W. Movie box office prediction based on multi-model ensembles. *Information*. 2022;13(6):299. DOI: 10.3390/info13060299
30. Lee S., Bikash K.C., Choeh J.Y. Comparing performance of ensemble methods in predicting movie box office revenue. *Heliyon*. 2020;6(6): e04260. DOI: 10.1016/j.heliyon.2020.e04260
31. Park J.H., Lim C. Predicting movie audience with stacked generalization by combining machine-learning algorithms. *Communications for Statistical Applications and Methods*. 2021;28(3):217–232. DOI: 10.29220/CSAM.2021.28.3.217
32. Dutta S., Dasgupta K. A shallow approach to gradient boosting (XGBoosts) for prediction of the box office revenue of a movie. In: Proc. Int. conf. on innovations in software architecture and computational systems. Singapore: Springer; 2021:207–219. DOI: 10.1007/978-981-16-4301-9_16
33. Su Y., Zhang Y., Yan J. Neural network based movie rating prediction. In: Proc. Int. conf. on Big Data and computing (ICBDC'18). (Shenzhen, April 28–30, 2018). New York, NY: Association for Computing Machinery; 2018:33–37. DOI: 10.1145/3220199.3220204
34. Ru Y., Li B., Chai J. An effective daily box office prediction model based on deep neural networks. *Cognitive Systems Research*. 2018;52:182–191. DOI: 10.1016/j.cogsys.2018.06.018
35. Wang W., Xiu J., Yang Z., Liu C. A deep learning model for predicting movie box office based on deep belief network. In: Tan Y., Shi Y., Tang Q., eds. Advances in swarm intelligence (ICSI 2018). Cham: Springer; 2018:530–541. (Lecture Notes in Computer Science. Vol. 10942). DOI: 10.1007/978-3-319-93818-9_51
36. Lu W., Zhang X., Zhan X. Movie box office prediction based on IFOA-GRNN. *Discrete Dynamics in Nature and Society*. 2022;3690077. DOI: 10.1155/2022/3690077

ИНФОРМАЦИЯ ОБ АВТОРЕ / ABOUT THE AUTHOR



Антон Валентинович Дождиков — кандидат политических наук, старший научный сотрудник кафедры ЮНЕСКО, Институт социально-политических исследований Федерального научно-исследовательского социологического центра Российской академии наук, Москва, Российская Федерация

Anton V. Dozhdikov — Cand. Sci. (Polit.), Senior Researcher at the UNESCO Chair, Institute of Socio-Political Research of the Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences, Moscow, Russian Federation
<http://orcid.org/0000-0002-1069-1648>
antondnn@yandex.ru

Конфликт интересов: автор заявляет об отсутствии конфликта интересов.
Conflicts of Interest Statement: The author has no conflicts of interest to declare.

*Статья поступила в редакцию 01.04.2024; после рецензирования 01.05.2024; принята к публикации 22.05.2024.
 Авторы прочитали и одобрили окончательный вариант рукописи.*

*The article was submitted on 01.04.2024; revised on 01.05.2024 and accepted for publication on 22.05.2024.
 The authors read and approved the final version of the manuscript.*

Переводчик Н.И. Соколова

Приложение А

**Результаты прогнозной классификации в соответствии с фактическими значениями /
Results of Predictive Classification in Accordance with Actual Values**

Название кинофильма / Title of the movie	2-классовая классификация / 2-class classification		4-классовая классификация / 4-class classification	
	Фактический класс / Actual class	Предсказанный класс / Predicted class	Фактический класс / Actual class	Предсказанный класс / Predicted class
Нахимовцы	0	0	0	0
Асфальтовое солнце	0	0	0	0
Развод. Фильм второй	0	0	0	0
Легенды «Орленка»	0	0	1	0
Архипелаг	0	0	0	0
Ночной режим. Фильм	0	0	0	0
Позывной «Барон»	0	0	0	0
Маленький принц	0	0	0	0
Лена и справедливость	0	0	1	0
Барбоскины Team	0	0	0	0
КАЛЕНДАРЬ МА(Й)Я	0	0	1	1
Турбозавры, вперед!	0	0	0	0
Преподы на всю голову	0	0	0	0
Там, где наш дом	0	0	0	0
Начать сначала	0	0	0	0
У самого Белого моря	0	0	0	0
Тибра	0	0	0	0
Недетский дом	0	0	0	0
Кэт	0	0	0	0
Лебединое озеро	0	0	0	0
Красная Шапочка	0	0	0	0
Кто там?	0	0	0	0
Родитель	0	0	0	0
Экспресс	0	0	1	1
Вместе веселее	0	0	0	0
Французский мастер	0	0	0	0
Далекие близкие	0	0	0	0
Мать моего сына	0	0	0	0

Продолжение приложения А / Appendix A (continued)

Название кинофильма / Title of the movie	2-классовая классификация / 2-class classification		4-классовая классификация / 4-class classification	
	Фактический класс / Actual class	Предсказанный класс / Predicted class	Фактический класс / Actual class	Предсказанный класс / Predicted class
Голиаф	0	0	0	0
Зина & Леха: Операция «Хвост и вымя»	1	1	3	3
Сердце Пармы	0	0	1	1
Сказка для старых	0	0	1	0
Лавстори	0	0	0	0
Иван Семёнов: Школьный переполох	1	1	2	3
Дополнительный урок	0	0	0	0
Плакать нельзя	0	0	0	0
Девушка на миллион	0	0	0	0
Грозный папа	0	0	0	0
Большое путешествие. Специальная доставка	0	0	0	0
Либерея: Охотники за сокровищами	0	0	0	1
Приключения маленького Бахи	0	0	0	0
Скоро кончится лето	0	0	0	0
Плотник	0	0	0	0
Велга	0	0	0	0
Петрополис	0	0	0	0
Кюннэй	1	1	3	3
Осанна	0	0	0	0
Заступница	0	0	0	0
Святитель	0	0	0	0
Я буду жить	0	0	0	0
Золотые соседи	0	0	0	0
Любовники	0	1	1	1
Петр I. Последний царь и первый император	0	0	1	1
Похожий человек	0	0	0	0
Звезды мне укажут путь	0	0	0	0
Волшебники	0	0	0	0
По-мужски	0	0	0	0
Проклятие. Мертвая земля	0	0	0	0

Продолжение приложения А / Appendix A (continued)

Название кинофильма / Title of the movie	2-классовая классификация / 2-class classification		4-классовая классификация / 4-class classification	
	Фактический класс / Actual class	Предсказанный класс / Predicted class	Фактический класс / Actual class	Предсказанный класс / Predicted class
Накануне	0	0	0	0
Земун	0	0	0	0
Ангелы милы	0	0	0	0
Тысяча дешевых зажигалок	0	0	0	0
Чинк: хвостатый детектив	0	0	0	0
Слон	0	0	0	0
Жанна	0	0	0	0
Я на перемотке!	0	0	0	0
Гриша Субботин	0	0	0	0
Царство против разбойников	0	0	0	0
Честный развод 2	0	0	0	0
Время года зима	0	0	0	0
Умная Маша	0	0	0	0
Сестры	0	0	0	0
Наша зима	0	0	0	0
Физфак	0	0	0	0
F20	0	0	0	0
Император	0	0	0	0
Пелевин	0	0	0	0
Елки 9	1	1	3	3
На ощупь	0	0	0	0
Омут	0	0	0	0
Тень. Взять Гордея	0	0	0	0
Щелкунчик и волшебная флейта	0	0	0	0
Тайный Санта	0	0	1	1
На выдохе	0	0	0	0
На тебе сошелся клином...	0	0	0	0
Балабанов. Колокольня. Реквием	0	0	0	0
Рядом	0	0	0	0
Маша и медведь в кино: 12 месяцев	0	0	0	0

Продолжение приложения A / Appendix A (continued)

Название кинофильма / Title of the movie	2-классовая классификация / 2-class classification		4-классовая классификация / 4-class classification	
	Фактический класс / Actual class	Предсказанный класс / Predicted class	Фактический класс / Actual class	Предсказанный класс / Predicted class
Непослушник 2	1	1	3	3
Бывшие. Happy end	0	0	0	0
Чук и Гек. Большое приключение	0	0	1	1
Мира	0	0	0	1
Елки-иголки	0	0	0	0
Чебурашка	1	1	3	3
Клипмейкеры	0	0	0	0
Умка	0	0	0	0
Молодость	0	0	0	0
Турбозавры. Зимние приключения	0	0	0	0
Путь за мечтой	0	0	0	0
Верблюжья дуга	0	0	0	0
Лили и море / grand marin	0	0	0	0
Высоцкий. Неизвестные страницы.	0	0	0	0
Не хороните меня без Ивана	0	0	0	0
Дорогие мама и папа	0	0	0	0
О чем говорят мужчины...	0	1	1	3
Один настоящий день	0	0	0	0
Лиза	0	0	0	0
Легкое знакомство	0	0	0	0
Белый ангел тундры	0	1	1	1
Свободные отношения	0	1	1	1
Быть	0	0	0	1
Снег, сестра и росомаха	0	0	0	0
Праведник	0	0	0	0
Непослушная	1	1	3	3
Снежная королева: разморозка	0	0	0	0
Эта любовь	0	0	0	0
Тембот	0	0	0	0
Женщины Алтая	0	0	0	0

Продолжение приложения А / Appendix A (continued)

Название кинофильма / Title of the movie	2-классовая классификация / 2-class classification		4-классовая классификация / 4-class classification	
	Фактический класс / Actual class	Предсказанный класс / Predicted class	Фактический класс / Actual class	Предсказанный класс / Predicted class
Бугун-былыр 2	1	1	2	2
Чекаго	0	0	0	0
Бешенство	0	0	0	0
Папы против мам	1	0	3	0
Женя и Ваня на краю света	0	0	0	0
Где проходит граница?	0	0	0	0
Нюрнберг	0	0	1	1
Здоровый человек	0	0	0	0
Беспринципные в деревне	0	0	0	0
Детектор	0	0	0	0
Поехавшая	1	1	2	3
Эскортница	0	0	0	0
Холодная гонка	0	0	0	0
Святой архипелаг	0	0	0	0
Медея	0	0	0	0
На солнце, вдоль рядов кукурузы	1	1	2	2
Почти как все	0	0	0	0
Сергий против нечисти. Шабаш	0	0	0	0
Кукольник	0	0	0	0
Спимговорим	0	0	0	0
Коты Эрмитажа	0	0	1	0
Айта	1	1	3	3
Любовь-морковь. Восстание машин	0	0	0	1
Боже! Чувствую приближение твое!	0	0	0	0
Странный дом	0	0	0	0
Наводнение	0	0	0	0
Я хочу! Я буду!	0	0	0	0
Прямой эфир	0	0	0	0
Волны	0	0	0	0
Доктор	0	0	0	0

Продолжение приложения А / Appendix A (continued)

Название кинофильма / Title of the movie	2-классовая классификация / 2-class classification		4-классовая классификация / 4-class classification	
	Фактический класс / Actual class	Предсказанный класс / Predicted class	Фактический класс / Actual class	Предсказанный класс / Predicted class
Зоськина заправка	0	0	0	0
Русский крест	0	0	0	0
Вызов	0	1	1	3
Яга и книга заклинаний	0	0	0	0
Помилование	0	0	0	0
Сокровища партизанского леса	0	0	0	0
Юра дворник	0	0	0	0
Меньшее зло	0	0	0	0
Маскерад	0	0	0	0
14+ продолжение	0	0	1	0
Булгаков. Этот мир – мой!	0	0	0	0
Хитровка. Знак четырех	0	0	0	0
Непосредственно Каха. Другой фильм	1	1	3	3
До рассвета	0	0	1	1
Маша и медведь в кино: скажите «ой!»	0	0	1	1
Узлы	0	0	0	0
Голова-жестянка	0	0	0	0
Солнце на вкус	0	0	0	0
Снегирь	0	0	0	0
Сквозь время	0	0	0	0
Микулай	0	0	0	0
Мальдивы подождут	0	0	0	0
Дыхание	0	0	0	1
Турбозавры: привет, сирена!	0	0	0	0
Статус	0	0	0	0
Джонни	0	0	0	0
Кентавр	0	0	1	1
Бизнес по-русски	0	0	0	0
За Палыча!	1	0	2	1
Баба Яга спасает мир	1	0	2	1

Окончание приложения A / Appendix A (continued)

Название кинофильма / Title of the movie	2-классовая классификация / 2-class classification		4-классовая классификация / 4-class classification	
	Фактический класс / Actual class	Предсказанный класс / Predicted class	Фактический класс / Actual class	Предсказанный класс / Predicted class
Свидетель	0	0	0	0
Как звезды	0	0	0	0
Квест	0	0	0	0
Руслан и Людмила. Больше, чем сказка	0	0	0	0
Чижик-прыжик возвращается	0	0	0	0
Девочка Нина и похитители пианино	0	0	0	0
Концерт отменяется	0	0	0	0

Источник / Source: составлено автором / Compiled by the author.

Appendix B

Фактические и предсказанные значения по сборам в прокате, просмотрам зрителей и рейтингу «Кинопоиска» / Actual and Predicted Figures for Box Office Receipts, Audience Views, and “Kinopoisk” Ratings

№	Название / Name	ID	Сборы факт. / Actual box office receipts	Сборы прогноз / Expected box office receipts	Просмотры факт. / Actual views	Просмотр прогноз / Predicted views	Рейтинг факт. / Actual rating	Рейтинг прогноз / Predicted rating
1	Нахимовцы	1403069	65 000 000	75 190 862	264 000	342 669	5,24	5,2
2	Асфальтовое солнце	1392757	9 100 000	14 684 488	36 000	58 174	7,10	6,78
3	Развод. Фильм второй	5042126	6 900 000	14 253 677	26 000	56 278	6,43	6,15
4	Легенды «Орлёнка»	4517052	62 000 000	44 973 131	284 000	169 044	7,47	7,66
5	Архипелаг	1256273	819 000	960 973	3 600	3 999	5,75	5,48
6	Ночной режим. Фильм	5034524	73 000 000	29 859 533	26 000	75 715	5,99	6
7	Позывной «Барон»	1395858	978 000	1 265 734	3 500	6 263	5,79	5,59
8	Маленький принц	572553	17 000	134 147	105	614	8,03	8,02
9	Лена и справедливость	4522328	82 000 000	49 468 637	30 000	119 982	6,55	6,72
10	Барбоскины team	4518842	97 000 000	113 138 269	448 000	616 840	8,11	7,99
11	Календарь ма(й)я	4963617	115 000 000	70 163 944	483 000	320 873	7,40	6,94
12	Турбозавры, вперёд!	5042646	17 000 000	27 931 832	95 000	104 736	7,60	6,36
13	Преподы на всю голову	5042247	11 000 000	6 346 878	4 600	15 196	4,50	5,31
14	Там, где наш дом	4934473	289 000	362 769	1 700	1 708	8,31	8,3

Продолжение приложения Б / Appendix B (continued)

№	Название / Name	ID	Сборы факт./ Actual box office receipts	Сборы прогноз / Expected box office receipts	Просмотры факт. / Actual views	Просмотр прогноз / Predicted views	Рейтинг факт. / Actual rating	Рейтинг прогноз / Predicted rating
15	Начать сначала	2000101	55 000 000	44 862 489	208 000	185 372	6,10	5,74
16	У самого Белого моря	4624744	568 000	751 443	4 000	3 683	6,42	6,39
17	Тибра	4518317	19 000 000	17 323 733	67 000	77 573	5,97	6,01
18	Недетский дом	4680791	81 000 000	29 539 678	29 000	120 898	6,53	5,82
19	Кэт	4680804	2 900 000	8 406 769	8 600	41 158	5,80	6,03
20	Лебединое озеро	1282100	24 000	1 406 730	81	4 571	6,25	5,96
21	Красная Шапочка	1445165	144 000 000	147 842 388	546 000	518 831	4,45	4,39
22	Кто там?	4999693	91 000 000	87 215 892	32 000	201 908	6,22	6,17
23	Родитель	1399249	20 000 000	5 417 875	6 900	21 429	7,21	7,39
24	Экспресс	2000127	81 000 000	37 800 832	27 000	108 361	6,37	6,47
25	Вместе веселее	5075871	51 000 000	20 383 758	9 500	71 274	5,05	5,19
26	Французский мастер	4528694	6 400 000	2 321 512	11 000	9 450	5,32	5,59
27	Далекие близкие	4489530	24 000 000	23 899 852	90 000	73 357	7,42	7,38
28	Мать моего сына	4400160	9 200 000	28 682 771	32 000	149 687	5,86	5,89
29	Голиаф	5047455	48 000	237 145	137	1 145	6,46	6,7
30	Зина & Лёха: Операция «Хвост и вымя»	5119760	15 000 000	11 095 392	46 000	48 230	Н/Д	4,31
31	Сердце Пармы	1045585	937 000 000	648 048 356	3 100 000	2 345 388	6,72	6,55
32	Сказка для старых	4739575	3 800 000	4 201 228	13 000	18 707	6,00	6,21
33	Лавстори	984365	70 000 000	61 204 656	253 000	234 582	4,68	5,63
34	Иван Семёнов: Школьный переполох	5021585	144 000 000	122 546 247	648 000	421 285	6,19	5,87
35	Дополнительный урок	1451292	12 000 000	9 790 986	43 000	46 126	6,98	6,51
36	Плакать нельзя	4322004	3 400 000	4 308 830	11 000	20 444	6,66	6,37
37	Девушка на миллион	5117212	132 000	239 639	608	1 136	4,89	4,48
38	Грозный пapa	1445164	158 000 000	108 267 371	662 000	478 651	6,01	5,74
39	Большое путешествие. Специальная ...	1445162	180 000 000	104 317 294	817 000	442 602	7,57	6,25
40	Либерея: Охотники за сокровищами	1398953	238 000 000	323 131 523	860 000	1 305 321	6,22	6,38
41	Приключения маленького Бахи	4530500	3 700 000	5 332 417	18 000	27 764	Н/Д	5,21
42	Скоро кончится лето	1443813	1 200 000	1 190 119	3 900	4 133	6,14	6,06
43	Плотник	4381764	1 400 000	11 233 990	4 100	56 437	6,61	6,77
44	Велга	1344504	825 000	1 079 702	3 000	9 060	6,28	5,96
45	Петрополис	1399052	2 700 000	4 828 853	9 100	29 594	5,42	5,8
46	Кюннэй	5160746	6 300 000	4 484 640	22 000	23 461	Н/Д	5,25
47	Осанна	4933779	88 000	857 851	518	8 312	8,50	7,29

Продолжение приложения Б / Appendix B (continued)

№	Название / Name	ID	Сборы факт./ Actual box office receipts	Сборы прогноз / Expected box office receipts	Просмотры факт. / Actual views	Просмотр прогноз / Predicted views	Рейтинг факт. / Actual rating	Рейтинг прогноз / Predicted rating
48	Заступница	1043835	182000	764612	1100	4438	Н/Д	6,08
49	Святитель	4959314	172000	932162	971	8734	8,46	7,17
50	Я буду жить	4660300	20000000	30428381	73000	113036	6,90	6,87
51	Золотые соседи	1328028	2400000	3599552	9100	13557	4,40	4,16
52	Любовники	4639557	176000000	163037230	586000	624977	5,90	6,05
53	Петр I. Последний царь и первый имп....	5116673	275000000	529076995	1200000	2323883	7,00	6,65
54	Похожий человек	5074909	554000	1759313	1700	8795	6,00	6,42
55	Звезды мне укажут путь	1162851	3400000	11182553	15000	56452	7,30	6,91
56	Волшебники	1008397	71000000	80602121	289000	392378	5,20	6,05
57	По-мужски	4559710	27000000	28027786	87000	110736	7,10	7,24
58	Проклятие. Мертвая земля	4837786	1200000	7014696	4200	29224	5,20	5,93
59	Накануне	4321128	841000	999946	3000	4658	5,40	5,07
60	Земун	4388687	3500000	4690601	12000	15947	6,20	6,21
61	Ангелы милы	4694772	196000	421344	868	1957	6,30	6,37
62	Тысяча дешевых зажигалок	5016804	1100000	1573941	3800	6038	6,30	6,28
63	Чинк: хвостатый детектив	5088766	85000000	91722969	354000	369629	6,20	6,12
64	Слон	5047298	20000000	21931224	68000	90386	6,60	6,48
65	Жанна	900055	3400000	57025433	11000	252381	6,70	6,27
66	Я на перемотке!	5126797	14000000	31093554	54000	110591	7,60	6,78
67	Гриша Субботин	1410951	4200000	5267243	17000	21104	5,20	4,96
68	Царство против разбойников	4853046	42000000	53329355	189000	252476	4,40	5,09
69	Честный развод 2	4856074	32000000	56615622	114000	194309	5,90	5,39
70	Время года зима	2000121	1800000	2838079	5500	10619	7,40	6,96
71	Умная Маша	4710763	15000000	25819110	57000	79271	5,90	6,14
72	Сестры	4661294	22000000	35564788	75000	121605	5,60	5,71
73	Наша зима	5104424	4600000	5732696	14000	25268	6,80	6,9
74	Физфак	5133122	259000	18388610	1200	148222	5,90	5,7
75	F20	5074907	480000	699433	1700	3363	5,70	5,57
76	Император	5066764	222000	251261	621	962	6,20	6,22
77	Пелевин	5195073	1300000	1216460	3400	4139	6,30	6,15
78	Елки 9	4958223	585000000	346912974	1900000	1514429	4,90	4,64
79	На ощупь	4975736	29000000	38459281	100000	138538	5,70	5,67
80	Омут	4913135	2900000	4488803	10000	16685	5,00	5,12
81	Тень. Взять Гордея	4625021	28000000	227049548	100000	869181	6,30	5,66

Продолжение приложения Б / Appendix B (continued)

№	Название / Name	ID	Сборы факт./ Actual box office receipts	Сборы прогноз / Expected box office receipts	Просмотры факт. / Actual views	Просмотр прогноз / Predicted views	Рейтинг факт. / Actual rating	Рейтинг прогноз / Predicted rating
82	Щелкунчик и волшебная флейта	1435400	115 000 000	112 315 705	483 000	438 012	7,90	7,07
83	Тайный Санта	4901797	84 000 000	64 549 229	297 000	290 577	5,70	5,22
84	На выдохе	4474120	927 000	1 064 510	3 300	4 363	6,00	6,04
85	На тебе сошелся клином белый свет	1392734	1 700 000	1 453 542	5 500	5 554	5,20	5,47
86	Балабанов. Колокольня. Реквием	5149154	2 400 000	3 414 803	7 000	24 239	8,10	7,6
87	Рядом	4819385	342 000	530 297	967	1 550	5,60	5,9
88	Маша и медведь в кино: 12 месяцев	5189169	109 000 000	123 711 842	499 000	461 413	8,50	8,03
89	Непослушник 2	5023667	518 000 000	204 013 845	1 700 000	720 589	6,10	6,39
90	Бывшие. Happy end	5091614	183 000 000	297 062 688	539 000	1 310 025	6,00	6,04
91	Чук и гек. Большое приключение	1346357	253 000 000	238 259 209	956 000	925 131	6,60	6,72
92	Мира	1227997	343 000 000	436 726 746	1 100 000	1 628 008	7,40	6,52
93	Елки-иголки	4694773	59 000 000	50 819 536	185 000	215 306	7,20	6,6
94	Чебурашка	4370148	7020 407 894	1 630 133 131	23 510 559	5 909 126	7,30	6,54
95	Клипмейкеры	4534033	22 000 000	18 306 527	65 000	100 910	6,00	6,15
96	Умка	5230220	9 900 000	17 223 971	50 000	91 658	6,00	6,35
97	Молодость	5074901	3 400 000	6 256 382	11 000	25 064	6,00	6,57
98	Турбозавры. Зимние приключения	5236779	21 000 000	23 833 876	108 000	110 362	6,00	6,21
99	Путь за мечтой	5079139	1 600 000	1 907 112	4 100	7 473	7,30	7,48
100	Верблюжья дуга	5129254	47 000	143 501	206	655	Н/Д	5,99
101	Лили и море / grand marin	4478721	1 200 000	2 773 415	3 400	8 571	6,20	6,31
102	Высоцкий. Неизвестные страницы. «од...	1101274	161 000	287 034	711	976	Н/Д	5,4
103	Не хороните меня без Ивана	5237750	10 000 000	11 652 379	34 000	52 359	8,30	7,78
104	Дорогие мама и папа	5249732	799 000	943 153	3 300	3 739	Н/Д	5,75
105	О чем говорят мужчины. Простые удо...	5045951	299 000 000	316 029 793	857 000	1 758 233	6,50	6,54
106	Один настоящий день	5047485	4 500 000	11 423 414	16 000	56 322	6,50	6,26
107	Лиза	1438922	2 600 000	3 189 975	11 000	13 730	5,80	5,76
108	Легкое знакомство	5245365	1 400 000	1 361 926	4 400	5 769	Н/Д	4,58
109	Белый ангел тундры	5254553	2 900 000	4 453 915	10 000	16 843	Н/Д	5,84

Продолжение приложения Б / Appendix B (continued)

№	Название / Name	ID	Сборы факт./ Actual box office receipts	Сборы прогноз / Expected box office receipts	Просмотры факт. / Actual views	Просмотр прогноз / Predicted views	Рейтинг факт. / Actual rating	Рейтинг прогноз / Predicted rating
110	Свободные отношения	4533254	158 000 000	209 632 007	510 000	776 981	6,90	7,04
111	Быть	4682864	123 000 000	114 574 015	436 000	489 336	5,50	5,58
112	Снег, сестра и росомаха	5129333	707 000	168 2918	1 700	8 202	6,30	6,27
113	Праведник	4484927	630 000 000	672 236 093	2 100 000	2 418 854	7,60	7,22
114	Непослушная	5189350	271 000 000	215 114 973	806 000	881 792	5,40	5,32
115	Снежная королева: разморозка	4756225	113 000 000	118 293 013	475 000	518 710	7,40	7,43
116	Эта любовь	5246539	1 900 000	1 812 341	6 500	7 420	5,90	5,87
117	Тембот	5034460	135 000	216 136	480	957	5,90	6,01
118	Женщины Алтая	5253702	61 000	145 889	218	656	Н/Д	5,89
119	Бугун-былыр 2	5078980	7 000 000	6 200 515	22 000	27 076	Н/Д	5,22
120	Чекаго	5077317	16 000 000	13 522 122	55 000	54 290	6,60	6,78
121	Бешенство	1346359	97 000 000	278 004 790	303 000	1 257 287	6,90	6,28
122	Папы против мам	4937051	119 000 000	86 542 200	427 000	346 151	4,50	4,22
123	Женя и Ваня на краю света	5034461	29 000	143 549	100	713	6,50	6,44
124	Где проходит граница?	5129306	3 400 000	3 633 843	26 000	26 229	Н/Д	5,66
125	Нюрнберг	905033	328 000 000	428 709 934	1 100 000	1 990 970	6,00	6,25
126	Здоровый человек	2000122	6 900 000	47 188 942	20 000	220 279	7,40	6,82
127	Беспринципные в деревне	5129252	312 000 000	294 513 410	876 000	1 198 275	6,50	6,34
128	Детектор	4694776	3 400 000	6 281 438	11 000	24 772	5,00	5,06
129	Поехавшая	4686066	478 000 000	420 677 635	1 600 000	1 748 092	7,80	7,33
130	Эскортница	5117304	55 000 000	50 478 520	161 000	223 909	4,60	5,12
131	Холодная гонка	5030860	18 000	126 849	83	577	5,20	5,07
132	Святой архипелаг	5101891	11 000 000	6 603 084	38 000	27 522	7,70	7,15
133	Медея	5089031	14 000 000	28 082 456	48 000	125 688	4,50	5,12
134	На солнце, вдоль рядов кукурузы	4729320	454 650 603	186 629 797	1 595 307	925 087	6,40	6,68
135	Почти как все	5078979	7 000 000	5 342 655	21 000	20 280	6,40	6,58
136	Сергий против нечисти. Шабаш	5126940	13 779 921	49 570 326	46 744	124 553	6,90	6,04
137	Кукольник	4528911	12 000 000	13 600 474	48 026	62 152	5,90	5,77
138	Спим говорим	5246830	12 000	531 577	62	6 167	5,80	5,44
139	Коты Эрмитажа	4907586	282 172 228	144 228 031	1 200 747	579 042	5,20	5,52
140	Айта	5104425	25 782 268	11 745 092	77 773	40 709	6,90	6,77
141	Любовь-морковь. Восстание машин	4715474	48 009 304	131 558 167	165 120	666 435	4,40	5,08
142	Боже! Чувствую приближение твое!	5266955	238 870	309 805	912	1 299	Н/Д	5,99

Продолжение приложения Б / Appendix B (continued)

№	Название / Name	ID	Сборы факт./ Actual box office receipts	Сборы прогноз / Expected box office receipts	Просмотры факт. / Actual views	Просмотр прогноз / Predicted views	Рейтинг факт. / Actual rating	Рейтинг прогноз / Predicted rating
143	Странный дом	5140353	18 000 000	12 574 717	61 000	58 847	4,90	4,57
144	Наводнение	5001203	218 000	1 443 540	557	7 049	5,60	5,84
145	Я хочу! Я буду!	4493006	20 396 022	3 403 6 615	75 750	117 771	6,30	5,79
146	Прямой эфир	4400163	881 000	27 460 151	2 634	153 902	6,80	6,2
147	Волны	5059142	82 000	1 711 084	245	10 090	5,40	4,61
148	Доктор	5002368	496 626	5 276 67	1 411	1 654	8,20	7,24
149	Зоськина заправка	5214412	992 542	22 542 776	3 347	116 461	5,40	5,2
150	Русский крест	4676639	14 266 106	25 192 400	58 000	95 270	7,50	7,58
151	Вызов	4448519	2 135 633 446	1 831 151 743	6 713 830	719 0055	7,60	7,27
152	Яга и книга заклинаний	5194326	224 991 304	179 621 842	943 294	730 191	7,80	6,76
153	Помилование	5255504	27 969 579	31 987 451	109 670	110 815	6,00	6,25
154	Сокровища партизанского леса	5264352	40 460 02	4 661 213	21 659	17 202	7,10	6,96
155	Юра дворник	5134108	51 876 565	52 737 549	177 188	162 620	6,80	6,81
156	Меньшее зло	4920455	155 872	26 033 180	501	111 844	7,00	6,99
157	Маскера	1320002	249 900	803 808	731	2 761	5,80	4,77
158	14+ продолжение	4489519	93 459 371	32 870 284	315 983	132 375	4,30	5,11
159	Булгаков. Этот мир – мой!	5277469	46 970	148 747	218	663	7,70	7,79
160	Хитровка. Знак четырех	5101673	105 619 225	66 927 173	361 396	281 500	7,00	6,41
161	Непосредственно Каха. Другой фильм	5236765	355 513 600	310 246 710	1 127 420	1 287 789	2,60	3,13
162	До рассвета	5254168	11 819 913	6 411 726	39 034	27 405	6,10	6,35
163	Маша и медведь в кино: скажите «ой!»	5264991	100 179 923	105 054 457	497 854	485 502	8,10	8,4
164	Узлы	5238327	162 086	256 413	559	1 008	5,80	5,81
165	Голова-жестянка	4493055	10 888 045	12 459 348	40 940	47 995	6,90	6,28
166	Солнце на вкус	5230216	28 469 452	37 316 323	144 074	167 361	7,10	6,94
167	Снегирь	4485219	56 233 634	54 201 899	185 902	229 464	7,10	7,17
168	Сквозь время	1045993	54 010 222	104 828 315	200 792	429 832	5,40	5,29
169	Микулай	5074908	2 377 184	2 156 652	8 107	9 878	6,60	6,6
170	Мальдивы подождут	5079082	57 138 515	53 508 509	189 574	209 671	6,40	5,47
171	Дыхание	4715481	114 702 994	204 653 433	403 389	928 421	6,60	6,16
172	Турбозавры: привет, сирена!	5278422	12 626 919	19 862 822	66 014	91 165	8,00	7,31
173	Статус	1346399	173 817	242 248	562	951	3,90	4,69
174	Джонни	5307214	124 132	260 285	440	962	6,40	6,34
175	Кентавр	5235968	123 095 316	95 594 774	400 593	405 597	7,50	6,83

Окончание приложения Б / Appendix B (continued)

№	Название / Name	ID	Сборы факт./ Actual box office receipts	Сборы прогноз / Expected box office receipts	Просмотры факт. / Actual views	Просмотр прогноз / Predicted views	Рейтинг факт. / Actual rating	Рейтинг прогноз / Predicted rating
176	Бизнес по-русски	5135024	9 600	128 667	46	617	5,30	5,19
177	За Палыча!	5235230	298 541 846	148 807 227	1 065 302	615 172	5,60	5,78
178	Баба Яга спасает мир	4536580	544 129 078	275 566 451	2 138 645	1 227 269	5,80	5,26
179	Свидетель	5332755	14 454 555	20 879 076	49 196	62 262	3,00	3,85
180	Как звезды	5313842	1 972 989	3 639 148	6 340	9 097	5,80	5,97
181	Квест	5118213	12 899 066	13 947 447	55 977	62 621	5,60	5,41
182	Руслан и Людмила. Больше, чем сказка	5356063	110 201 546	133 865 184	459 620	568 500	8,20	7,48
183	Чижик-Пыжик возвращается	5003770	40 760 444	38 860 073	182 266	167 487	7,30	7,07
184	Девочка Нина и похитители пианино	4667350	13 150 750	26 628 390	57 799	94 541	5,40	4,97
185	Концерт отменяется	5325618	462 037	1 126 690	1 460	5 357	5,70	5,9
Итого за период:			21702451281	15374212770	73465643	62363992		

Источник / Source: составлено автором / Compiled by the author.

Комментарий: данные представлены по состоянию на 1 декабря 2023 г. Количество просмотров и величина сборов может измениться в большую сторону, зрительский рейтинг – в большую и меньшую стороны.

Аномалией, показывающей работу административного ресурса, подбора времени проката и исключение конкурирующих кино- картин, является фильм «Чебурашка», который при всех прочих условиях окупился бы в прокате и получил хорошие показатели, но с учетом выхода в новогоднее время, при отсутствии конкуренции и активной рекламной кампании он показал высокий уровень сборов.